Current Index to Statistics

User Guide
CIS EXTENDED DATABASE

1997 Edition

Sponsored by
The American Statistical Association and
The Institute of Mathematical Statistics

Ronald A. Thisted
Department of Statistics, The University of Chicago

Michael J. Wichura
Department of Statistics, The University of Chicago

Klaus Hinkelmann
Department of Statistics, Virginia Tech

Bruce E. Trumbo
Department of Statistics, California State University, Hayward

November 1997

Editor, *Current Index to Statistics*

Ronald A. Thisted	1995–

Editors, *CIS* Printed Volume

Klaus Hinkelmann	Volumes 21–	1995–
Edward E. Gbur, Jr.	Volumes 16–20	1990–94
Richard K. Burdick	Volumes 11–15	1985–89
James E. Gentle	Volumes 6–10	1980–84
Brian L. Joiner	Volumes 1–5	1975–79

Editors, *CIS/ED*

Michael J. Wichura	1995–
Ronald A. Thisted	1994–95
Bruce E. Trumbo	1990–93

CIS Management Committee, 1997

Paul Shaman, Chair

Douglas M. Bates	Michael J. Meyer
Julia A. Norton	Alan M. Zaslavsky

ISBN 1–883276–52–7
ISSN 1094–7469

Copyright © 1997
American Statistical Association & Institute of Mathematical Statistics
All rights reserved.

Licensees of the 1997 *CIS/ED* are granted a non-exclusive, royalty-free license to copy and to distribute this *User Guide* to authorized users of the database. Each such copy must contain the title page and this page in its entirety.

No bibliography, index, or other work (whether in computer-readable, printed or other form) derived from more than 250 Database records contained in the Database may be distributed without the written permission of the authorized representative of the *CIS/ED*. Any publication prepared with the substantial assistance of the Database should acknowledge its use and cite it as a reference: "Michael J. Wichura, Editor (1997). Current Index to Statistics Extended Database. American Statistical Association (Alexandria, VA) and Institute of Mathematical Statistics (Hayward, CA)."

Certain programs, files, and documents on CD-ROM distribution of the *CIS/ED* are copyright by their individual authors. Each such item contains a notice of copyright, and may also contain other restrictions on redistribution described herein.

Information concerning current prices, licensing, and availability can be obtained from our page on the World Wide Web, (URL: `http://www.stat.uchicago.edu/~cis`), or by contacting either the ASA (+1 703 684 1221) or IMS (+1 510 783 8141) offices.

Contents

I Quick Start 1

1 Quick Guide to Installation 1

1.1 All Installations 1
1.2 Installing CIScd Search Software 2
1.3 Hard Drive Installation on Desktop Computers 4
1.4 Unix™ CD-ROM Installation 4
1.5 Unix™ Hard Drive Installation 5
1.6 Technical Support 5

II Search Software 6

2 CIScd, Version 1.7.2 6

2.1 Installation 6
2.2 How the program works 7
2.3 Tutorial: Getting Started 8
2.4 An Initial Refinement 9
2.5 Restricting the Search 9
2.6 A More Difficult Example 10
2.7 Commands 13
2.8 Search Commands 14
2.9 Output formats 18
2.10 Setting output format 19
2.11 Saving to a file (`save`) 19
2.12 Loading a saved file (`retrieve`) 20
2.13 More (`more`) 20
2.14 Screen display (`list`, `display`) 20
2.15 Showing the next k records (`next`) 20
2.16 Manipulating the search-result stack 21
2.17 Displaying the contents of a disk file (`type`) 21
2.18 Displaying settings 21
2.19 Version 21
2.20 Comment (`#`) 21
2.21 Looking up an abbreviation (`whatis`) 21

2.22 Using CIScd with a printer (DOS/WINDOWS only) . . . 22
2.23 Restarting, or changing configurations . . . 22
2.24 Quit (quit) . . . 22
2.25 Modifying the cis.cfg when Files are on Hard Disk . . . 23
2.26 Fine Print . . . 23

III User Reference 24

3 Overview 24
3.1 Organization and Contents . . . 24
3.2 Search Software. . . . 26
3.3 User Guide . . . 26
3.4 Technical Reference Manual . . . 27
3.5 Changes in Future Editions . . . 27

4 The Database Records 29
4.1 Field 1: Location . . . 29
4.2 Field 2: Title . . . 30
4.3 Field 3: Authors . . . 31
4.4 Field 4: Key words and phrases . . . 33
4.5 Field 5: Searchable author information . . . 33
4.6 Field 6: Additional Searchable Information . . . 34
4.7 Field 7: Administrative codes . . . 34

5 Hints for Searching CIS/ED 35

6 Source Abbreviations 39

7 Special Indexes 40
7.1 Subject Index . . . 40
7.2 Author Index . . . 42
7.3 Index of Key Words and Phrases . . . 45
7.4 Index of Additional Searchable Terms . . . 45
7.5 Inverted Indexes . . . 46
7.6 Thesaurus . . . 47

8 Abstracts and Book Reviews 48

IV Coverage **50**

9 The Gap Project **50**
9.1 Editorial policy 50
9.2 Contents 51
9.3 General Acknowledgments 53

10 License Information and Legal Matters **54**

11 History of the *CIS/ED* **56**
11.1 Information from printed volumes of CIS 56
11.2 Acknowledgements 57

12 Errors, Omissions, and Corrections **58**

Part I
Quick Start

1 Quick Guide to Installation

Read this chapter of the *User Guide* before installing the database.

1.1 All Installations

Last-minute Changes

Before beginning installation, please be sure to read the file `README.TXT` *at the top level of the CD-ROM volume. This file will contain information about any last-minute changes in procedure that differ from those described here.*

For More Information

Information, latest versions of search software, corrections, and announcements concerning the *Current Index to Statistics* are posted regularly on the World Wide Web. The *CIS* home page can be found by pointing your Web browser to `http://www.stat.uchicago.edu/~cis`.

Updating an Earlier Version of *CIS/ED*

Previous editions of the *CIS/ED* must be *removed* before installing the 1997 edition.

Do not attempt to "update" the 1996 edition of the CIS Extended Database by just adding the additional years from the 1997 edition of *CIS/ED*. Doing so will corrupt the database and will produce misleading or erroneous results.

Your license agreement prohibits mixing editions in the same installation. The main reason for this is that the additions that each new printed *CIS* volume make to the database extend far beyond publications in the most recent year. Approximately 25% of the entries in an annual printed volume carry earlier nominal publication dates. In fact, *CIS* editors are constantly trying to pick up items overlooked in previous years. As we

prepare each edition of *CIS/ED*, these additions will be sorted according to the years of publication of items indexed.

In general, there are other good reasons not to mix versions. The ongoing effort to attach correction notes and comments to their referent papers creates a subject-matter incompatibility with earlier editions. Changes and additions are made to the source abbreviations each year. Ongoing structural changes may make editions incompatible with each other. Finally, the inverted indexes distributed with the database will not work with mixed versions.

1.2 Installing CIScd Search Software

The steps to follow for installing search software depend on the computer system that you are using. For desktop computers, the easiest way to use the database is to install the search program CIScd on your hard drive, and to use it to access the database directly from the CD-ROM.

Instructions for installing this program on Macintosh, MS-DOS, and Windows computers follow. Once the software is installed, refer to Chapter 2 for full documentation.

Macintosh Installation

1. Drag the *Current Index Software for Mac* folder onto your hard drive. This folder may be left permanently on the hard drive.

2. Open the folder *on your hard disk* that you just created. [The program will not run if it is launched from the CD-ROM.]

3. Read the instructions for CIScd found in Chapter 2.

4. Double-click on the CIScd icon and start searching.

The CIScd application in the *Current Index Software for Mac* folder is a "fat binary" which will work on computers running the MacOS on either PowerPC or 680x0 platforms. The folder MACSFT contains versions of CIScd which are PPC- or 68k-specific. These versions can simply replace the copy of CIScd in the folder on your hard disk.

MS-DOS Installation

1. Insert the *CIS/ED* CD-ROM into the CD-ROM drive. The drive letter assigned to the CD-ROM differs from one computer to the next. In this example, we shall assume that the CD-ROM is in drive `F:`, and that your hard disk is drive `C:`.

2. Give the command `mkdir C:\DOSSFT`

3. Give the command `XCOPY F:DOSSFT C:\DOSSFT /S`

4. Give the command `CD C:\DOSSFT`

5. Give the command `INSTALL`

6. When prompted, give the drive letters that correspond to your CD-ROM drive. This program will create a file called `CIS.CFG` in the `DOSSFT` folder that tells the search program how to find the information on the CD-ROM in your computer.

7. Read the instructions for CIScd found in Chapter 2.

8. To use the CIScd search program, make certain that you are in the `DOSSFT` directory on your hard drive, and give the command `CIScd`. Note: the program will *not* run properly if you run the copy of `CISCD.EXE` directly from the CD-ROM itself.

Windows Installation

The easiest way to use the *CIS/ED* on computer running Microsoft Windows is to use the search program `CISCD.EXE` from an MS-DOS window, and to access the database directly from the CD-ROM.

Windows 3.1

From the File Manager, open the icon "MS-DOS Prompt," then follow the instructions given under MS-DOS installation in the previous section.

Windows 95

1. In the Windows Explorer, drag the `DOSSFT` folder from the CDROM to your hard disk. This will copy that folder and its contents to your hard disk.

2. While still in Windows Explorer, double-click on the `DOSSFT` folder.

3. While still in Windows Explorer, double-click on the `INSTALL` application.

4. When prompted, give the drive letter that corresponds to your CD-ROM drive. This program creates a file called CIS.CFG in the `DOSSFT` program on the hard disk. This file tells the search program how to find the information on the CD-ROM on your particular computer. (When the `INSTALL` application finished, it will print "All Done" on the screen in an MS-DOS window. You will need to close the MS-DOS window that the program opened yourself.)

5. Read the instructions for CIScd found in Chapter 2.

6. To use the CIScd search program, make sure that the CDROM is in the drive, and then double-click on the CIScd application in the `DOSSFT` folder on your *hard disk*. Note: the program will *not* run properly if you run the copy of `CISCD.EXE` directly from the CD-ROM itself.

   When you quite the CIScd application, you will need to close the MS-DOS window that it opened manually.

1.3 Hard Drive Installation on Desktop Computers

If you have the space to spare on your hard disk (about 40Mb), access to the database can be speeded considerably by copying the contents of the CD-ROM to your hard disk. Consult the documentation for CIScd (see also Part II. "Search Software") for information about setting up the configuration files in this case.

1.4 Unix™ CD-ROM Installation

To use the CD-ROM directly, mount the CD-ROM (typically using the `mount` or `cdmount` commands). On some SUN systems it is critical to inform the system that the CD-ROM is in "High Sierra" format (`-hsfs`) and is not a Network File System (`-nfs`). If the device is, say, `/cdrom`, then all files and documents are contained in the directory `/cdrom/unix`. Database records are in the directory `/cdrom/unix/files`; index records are in the directory `/cdrom/unix/index`; and abstracts and book reviews are in the directory `/cdrom/unix/abstract`.

1.5 Unix™ Hard Drive Installation

To use a hard disk rather than the CD-ROM, mount the CD and copy the `/cdrom/unix` directory to the hard disk. Only the subdirectories of `unix` that will actually be used need to be copied from the CD-ROM. For instance, if the search software being used does not employ the inverted indexes, you may copy just the author, subject, and keyword index files (`author.v97`, `subject.v97`, `keywd.v97`, `f6.v97`) from the `index` subdirectory (or you may wish to omit this directory entirely). Be certain to configure any search software to look for the database in the new hard-disk directory.

Several options are available for search software on Unix™ platforms. Several contributed packages are contained and documented in the directory `/cdrom/unixsft`.

1.6 Technical Support

Additional information on special installations and corrections/additions may be found at the *CIS* Web page `http://www.stat.uchicago.edu/~cis` .

Part II
Search Software

2 CIScd, Version 1.7.2

CIScd is a program for searching the *Current Index to Statistics* Extended Database (*CIS/ED*) for entries that contain combinations of keywords specified by the user. Versions of CIScd that run under MSDOS, Windows, and MacOS operating systems are included on the CD-ROM.

This program is one of several that have been contributed by the *CIS/ED* user community; other programs, which run under Unix and other operating systems, can be found on the CD-ROM or via the *CIS* web site.

CIScd is not fancy, but it can be used to conduct quite intricate searches. Although it does not currently have a menu-driven interface, much of its power can be realized by learning a handful of commands.

2.1 Installation

The program is distributed in both Macintosh and MSDOS formats. The descriptions below are written for the Macintosh version, except where DOS-specific instructions are required, in which case the DOS versions appear in [square brackets].

The first step is to copy CIScd to a writable storage medium such as a hard disk. There must be enough free space on this disk to accomodate the results of searches and intermediate results. Except for the most extensive searches, 1Mb of space will suffice, and for most purposes, 500 Kb works well. (A floppy disk will work, but the diskette will be much slower and will be unable to accomodate the results of large searches.)

In order to operate properly both the application itself [`CISCD.EXE` on DOS systems] and the file `CIS.CFG` must be present in the same folder [directory] on the hard drive.The version of `CIS.CFG` supplied on the distribution diskette will read data and indexes from the CD-ROM. [On DOS machines, an installation program must be run to create the configuration file, as described in section 1 of this manual.] If you search databases located on different drives, you may have a separate configuration file for each setup and then switch between them as needed.

2.2 How the program works

To use the search program to its best advantage, it is essential to learn how the program works. The Current Index Database on CD-ROM actually consists of a set of files containing the *CIS* entries themselves, as well as a set of "index files" which are nearly as large. These index files are tables of keywords together with the location of records in the *CIS* file that contain those keywords. The keywords used to construct the index files on the CD-ROM were all words that

1. begin with an alphabetic character,
2. are followed by a numeral, punctuation mark, or blank space,
3. contain at least two characters, and
4. are no more than six characters in length.

Words in the database that exceed six characters are, for purposes of indexing, truncated to six characters. Thus, the words "regress", "regression", and "regressor" are all considered instances of the index word "regres"'. For purposes of indexing, upper and lower case letters are considered equivalent. Certain very common words (such as "of", "by", "and", etc) are also not indexed.

The author, title, keyword, and alternative spelling fields of the records are indexed.

A *search* looks up one or more keywords in the index files, and then uses the list of *CIS/ED* entries found there to construct a new *search result* containing all of the entries that match the keyword criteria.

Once a search result is obtained, the search can then be *refined* by adding additional constraints, or by combining the current search result with one obtained earlier.

CIScd operates on a *stack* of search results (in a fashion similar to calculators that use reverse-Polish notation). Each new `search` command creates a new search result, which becomes the top-most search result on the stack. All previous search results are "pushed down" onto the stack, maintaining the order in which they were created. Refinement commands operate either on the most recent search (top of the stack), or the two most recent searches. The contents of the stack after executing a command depend on the particular operation being carried out, but generally, the top-most search result on the stack is the output from the last operation.

Always remember: **Search** first, then **refine.**

2.3 Tutorial: Getting Started

To start a search, open (or double-click on) the CIScd icon — the one on your hard disk, not the CD-ROM. [DOS: give the command `CISCD`.] Something like the following will appear on the screen. The examples in this chapter use the *1996 Edition* of the database. Output using later versions of either the database or of CIScd will of course be slightly different.

```
Current Index to Statistics / Extended Database Search Software

Version 1.7.2 8-Mar-97 Ronald Thisted [94-95 file format]

Copyright 1995-7 Ronald Thisted
Portions of Search Engine Copyright 1993 Douglas M. Bates
All rights reserved.

CIS Extended Data Base
Publication or other use of more than 250 records
requires permission of the CIS.

== Type  'help'  for help.

=>
```

To search for all papers by William Krasker, enter "search Krasker" at the prompt. This will produce

```
=> search Krasker
...........
[16 records found]
[16 main records/0 backup records]
```

The dots are produced sequentially as the CD-ROM is searched. The time required for this query to find these 16 of the 150,000+ records on a Power Macintosh was about one second.

An alternative would be to enter "`search William Krasker`" at the prompt. In that case, all records that contain both `William` and `Krasker` will be produced. Note that the software automatically converts "William" to "willia" and "Krasker" to "kraske" before doing the search in the index. This search takes a bit longer (about 4 seconds), since every record that contains "William" must be checked to see if it also contains "Krasker".

2.4 An Initial Refinement

Having done the search above we can *refine* the results using any of a collection of powerful commands. These commands can do arbitrary Boolean searches on terms (or even on strings of consecutive characters). The `16 main records` found in the original search are stored in a temporary file on your disk.

The search refinement commands do not operate on the database as a whole, but only on this subset of records contained in the "main records" file. Each command that operates on the "main records" file leaves its result in the "main records", and the previous result in the "backup records".

In fact, CIScd creates a stack of search results. This stack has two visible elements, the "main results" and "backup results" file. This stack can be manipulated with commands such as `push`, `pop`, and `swap`. The file operations can be thought of commands given to a reverse-Polish calculator. The `stack` command gives a status report of all files in the stack. The `push` and `pop` commands are used to manipulate the stack.

These commands tend to be quite fast. For instance, starting with the 16 "Krasker" records, we can issue the request "and William", which produces

```
=> and William
[13 (of 16) records meet criteria]
[13 main records/16 backup records]
```

The 13 remaining records now contain both "William" and "Krasker" somewhere in them. This search was virtually instantaneous.

What were the characteristics of the 3 records omitted in the refinement? Two of them were by others commenting on a procedure of Krasker and Welsch. The other was by W. S. Krasker and J. W. Pratt — an article that we would like to have found! The difficulty is that names of authors are rarely consistent across journals or other publications, and often authors' names are misspelled. (The author index on the *CIS* CD-ROM can be browsed to find similar names represented on the database.)

2.5 Restricting the Search

Here is an alternative strategy for refining the initial search—use a search restricted to the `author` field of the records. After the initial search on "Krasker", we could proceed with

```
=> search krasker
=> author krasker
[14 (of 16) records meet criteria]
```

which finds the appropriate entries.

2.6 A More Difficult Example

How should we find articles written by Al Best? Here is a transcript that illustrates several features of CIScd.

```
=> #
=> # First, do an initial SEARCH (which uses the inverted index)
=> #
=> search best
.............................
[836 records found]
[836 main records/13 backup records]
=> #
=> # That took over a minute, due to the number of records
=> # that had to be retrieved.  A moment's reflection suggests
=> # that many of these records will deal with ``Best invariant
=> # tests'', or ``Best linear''.  So we refine the search by
=> # looking only in the author field.
=> #
=> author best
[64 (of 836) records meet criteria]
[64 main records/836 backup records]
=> list

71Biomtrcs 27     895- 901 J
Descriptive functions in disease
Best, William R.;Becktel, Jack M.;Johnson, Arthur F.
Regression;Discriminant analysis;Public health;Multivariate analysis

73Biomtrka 60     429- 430 J
Extended tables for Kendall's tau
Best, D. J.
Rank correlation coefficient;Nonparametric regression;Ties in ranks

74ApplStat 23      98- 100 J
[Algorithm AS 71] The upper tail probabilities of Kendall's tau
```

```
Best, D. J.;Gipps, P. G.
Edgeworth approximation

74Biomtrka 61     385- 386 J
The variance of the inverse binomial estimator
Best, D. J.
Proportion defective;Inverse binomial;Mean squared error;Sequential
sampling

74Technmcs 16     621- 624 J
-- MORE? [ynq] q
=> #
=> # Although Best is in the author field, many of them are not
=> # the individual we are after.  We can search for records
=> # whose author fields contain ‘‘A.’’ and ‘‘Best"
=> allauthor A. Best
[5 (of 64) records meet criteria]
[5 main records/64 backup records]
=> list

83Biomtrka 70     447- 453 J
A test for comparing large sets of tau values
Best, D. J.;Cameron, M. A.;Eagleson, G. K.
Kendall’s tau;Multiple comparisons;Poisson-limit theorem

83STATCOMP83       41-  49 P
Power approximations for Pearson’s chi-squared test
Best, D. J.;Rayner, J. C. W.;Turnbull, A. P.
Goodness-of-fit

89AstrlJSt 31     491- 492 J
Review of ‘‘Goodness-of-fit statistics for discrete multivariate data’’
Read, Timothy R. C.;Cressie, Noel A. C. (Auth);Best, D. J. (Rev)

91Biomtrcs 47     788- 789 J
Review of ‘‘Smooth tests of goodness of fit’’
Rayner, J. C. W.;Best, D. J. (Auth);Kemp, A. W. (Rev)

93JRSS-B   55      39-  52 J
Modelling complexity: Applications of Gibbs sampling in medicine (Disc:
p53-102)
Gilks, W. R.;Clayton, D. G.;Spiegelhalter, D. J.;Best, N. G.;McNeil, A.
```

```
J.;Sharples, L. D.;Kirby, A. J.
=> #
=> # Well, that was positively unhelpful -- none were the targets
=> # of our search!  We can back out of the last result by
=> # interchanging (SWAPping) the main and backup files:
=> swap
[64 main records/5 backup records]
=> #
=> # The ALLAUTHOR command finds records that contain ALL of the
=> # listed search terms in the author field.  We can also request
=> # records that meet ANY of a set of criteria, using the AUTHOR
=> # command -- Think of AUTHOR = AUTH + OR:
=> author Al Alvin Theodore Simon
[8 (of 64) records meet criteria]
[8 main records/64 backup records]
=> list

79ASAProStCp      153- 153 P
Discussion of LINMOD: A flexible system for multivariate linear models
computations
Best, Alvin M., III

79Psymtrka 44     395- 408 J
On the precision of a Euclidean structure
Best, Alvin M., III;Young, Forrest W.;Hall, Robert G.
Multidimensional scaling;Data analysis

   (etc)

93JRSS-B   55      39-  52 J
Modelling complexity: Applications of Gibbs sampling in medicine (Disc:
p53-102)
Gilks, W. R.;Clayton, D. G.;Spiegelhalter, D. J.;Best, N. G.;McNeil, A.
J.;Sharples, L. D.;Kirby, A. J.
=> #
=> # Near Success!
=> # A better way to accomplish the last search is to conduct a
=> # search on the string of characters ‘‘Best, A’’ in the
=> # author field:
=> swap
[61 main records/8 backup records]
=> author "Best, A"
```

```
[7 (of 64) records meet criteria]
[7 main records/64 backup records]
=> #
=> # Real Success!!
=> #
=> # We can save these results to a file, in exactly the format above:
=> #
=> save best.cis
[7 records written to file best.cis in SHORT format]
=> quit
```

If we suddenly were to discover that we forgot something, we can start CIScd again, and the main and backup results will be exactly as we left them, since the temporary files are not erased at the end of a session. If you would like to erase them (if there are some big search results, for instance, that you won't need again), you can add the word "clear" to the end of the quit request.

2.7 Commands

A list of commands can be obtained at any time with the "`help`" request. This produces a list like this:

```
help     Prints this text
swap     Interchange main and backup searches
count    Show current contents of main and backup searches
config   Initialize and restart, with optional config file
and      Finds records in current search with ALL keys
andor    Finds records in current search with ANY keys
not      Finds records in current search matching NO keys
notboth  Omits only records that have ALL specified keys
merge    Merges current and backup search, REPLACING current
save     Saves results of current search to named file
retrieve Retrieves a named file of CIS database records
display  Show results of main search in current format
quit     Quit current program
set      Set options.  For more info, type 'HELP SET'
show     Show options.  For more info, type 'HELP SET'
batch    Read search commands from named file
search   Extract records from CISED using inverted index
version  Display version number of software and database
more     Set pager options
```

```
#         Comment
author    Find records with ANY of keys in author field only
allauth   Find records with ALL of keys in author field only
noauth    Omits records with ANY of keys in author field only
type      Display the contents of a disk file to the screen
next      Display the next k records. Use 'NEXT k' to set k
diff      Find all records on backup but not on main search
whatis    Looks up an abbreviation in the abbreviation file
```

2.8 Search Commands

2.8.1 The preliminary SEARCH

Every *CIS* search will start with a broad search using the inverted index to the database. The SEARCH command looks for all entries in the database that match the first six (case insensitive) characters of each of the keywords listed on the request. Records that match all of the specified keywords are retained as the main search result. If at least one record was found, the previous main search result becomes the current backup search result.

2.8.2 Refinements

The syntax for the requests that refine searches is more flexible. Each request has the form

```
REQUEST  Token1 Token2 ... TokenK
```

where REQUEST represents a command such as NOT, AND, ANDOR, AUTHOR, or ALLAUTH. If the first four characters of REQUEST match a valid request, the corresponding request will be honored. Thus, AUTH is an abreviation for AUTHOR.

Each token consists either of a string of non-blank characters, or a string of characters that include blanks, enclosed in double quotation marks ("). Neither the double quotation mark (") nor the tilde (~) is used in the *CIS*, and these characters are not valid in search tokens. The number of tokens on a single request is not limited (for practical purposes).

After each successful refinement request, the previous master search result becomes the current backup, and the results of the request become the current master.

Here are the common refinement commands, and a brief description of their function:

```
and      Finds records in current search that have ALL search terms
andor    Finds records in current search that have ANY search terms
not      Finds records in current search that match NO search terms
notboth  Omits records having ALL search terms, keeps all others
merge    Merges current and backup search, REPLACING current search
author   Find records with ANY of keys, in author field only
allauth  Find records with ALL of keys, in author field only
noauth   Omits records with ANY of keys, in author field only
diff     Find all records on backup but not on main search
```

2.8.3 Example

Papers which cite, refer to, or are written by (any) Efron, and which cite, refer to, or are written by (any) Stein:

```
=> search efron
.............................
[138 records found]
[138 main records/7 backup records]
=> and stein
[18 (of 138) records meet criteria]
[18 main records/138 backup records]
```

Of these papers, some are written by Efron and Stein; ALLAUTH searches for records whose author fields contain all of the listed tokens.

```
=> allauth efron stein
[1 (of 18) records meet criteria]
[1 main records/18 backup records]
=> list

81AnlsStat  9     586- 596 J
The jackknife estimate of variance
Efron, B.;Stein, C.
Analysis of variance;Bootstrap;$U$-statistic
94x;U-statistic
```

Of course, some of the papers might have been written by Efron about Stein or his work; after returning to the original search results (using SWAP) we have

```
=> swap
[18 main records/1 backup records]
```

```
=> author Efron
[12 (of 18) records meet criteria]
[12 main records/18 backup records]
=> list

71JASA     66     807- 815 J
Limiting the risk of Bayes and empirical Bayes estimators -- Part I:
The Bayes case (Ref: V67 p130-139)
Efron, Bradley;Morris, Carl
Stein estimator;Shrinkage estimator
...
```

These 12 references also contain the Efron/Stein paper cited earlier. To find the papers written by Efron and Morris regarding Stein shrinkage estimation, we use:

```
=> allauth efron morris
[9 (of 12) records meet criteria]
[9 main records/11 backup records]
```

Of the original 18 papers (with two `pop`'s from this point), how many dealt with the "Efron-Stein inequality"?

```
=> pop
[12 main records/18 backup records]
=> pop
[18 main records/1 backup records]
=> count
[18 main records/0 backup records]
=> and Efron-Stein
[4 (of 18) records meet criteria]
[4 main records/18 backup records]
=> list

  84IneqlSt&Pr   112- 114
  An expansion for symmetric statistics and the Efron-Stein inequality
  Richard A. Vitale
  U-statistic

  86AnlsStat14   753- 758
  An Efron-Stein inequality for nonsymmetric statistics
  J. Michael Steele
  Symmetric function;Variance bound
```

```
  87AnlsStat15  1317-1320
  An application of the Efron-Stein inequality in density estimation
  Luc Devroye
  Kernel estimator

  88StPrLet  7   105- 112
  A differential version of the Efron-Stein inequality: Bounding the
variance of a function of an infinitely divisible variable
  Richard A. Vitale
```

Note that the `SEARCH` command would not accept "Efron-Stein" as a search token — or more accurately, it would take the hyphen as a word-boundary, and search for the token "efron", which we know produces more than 100 matches. Similarly, we could have refined the search in the following way (note the use of quotation marks):

```
=> and ‘‘Efron-Stein inequality"
[4 (of 18) records meet criteria]
```

2.8.4 Another Example

Here is an example of commands that find all indexed works by Paul Switzer written without a co-author (and excluding reviews):

```
=> search switzer
..................
[39 records found]
[39 main records/4 backup records]
=> author Switzer
[33 (of 39) records meet criteria]
[33 main records/39 backup records]
=> #
=> # That gets rid of 'Switzerland', for example
=> noauth (Rev)
[27 (of 33) records meet criteria]
[27 main records/33 backup records]
=> # Now the reviews written by Switzer are gone
=> noauth ;
[14 (of 27) records meet criteria]
[14 main records/27 backup records]
=> # That deleted all multi-authored works (!)
```

```
=> author "Switzer, P"
[14 (of 14) records meet criteria]
[14 main records/14 backup records]
=> # The last request verifies that the author field of all of
=> # the selected records contain both 'Switzer' as a last name,
=> # with a given name that starts with P (such as 'P.' or 'Paul')
=> pop
[14 main records/27 backup records]
=> # This command "pops" back up one step
=> diff
[13 records meet criteria]
[13 main records/14 backup records]
=> # These should contain only jointly-authored papers
=> author "Switzer, P"
[8 (of 13) records meet criteria]
[8 main records/13 backup records]
=> # These papers contain an author with one of the variants above:
=> list

82JIAMaGeo 14     433- 444 J
A prior probability method for smoothing discriminant analysis
classification maps
Switzer, Paul;Kowalik, William S.;Lyon, Ronald J. P.
[etc]
=> whatis JIAMaGeo
   Abbreviation 'JIAMaGeo' corresponds to
   Journal of the International Association for Mathematical Geology
```

2.9 Output formats

The search results (main and backup) are stored using the same cryptic format that the Current Index Database itself uses; we call this format the **raw** format. By default, requests to list or display the current search will produce screen output in an abbreviated, human-readable format called **short** format. A longer format (**long**) lists the contents of each fields on a separate line.

Refer format outputs each record in a form that can be used by the UNIX bibliographic utility *Refer*, as well as by some other bibliographic management software (such as *EndNote* on the Macintosh).

Tex format produces a file that can be run through the TEX typesetting language to produce reasonably readable bibliography lists. These may require some hand editing.

2.10 Setting output format

There are two ways to change the output format from the default value of **short**. One option is to permanently change the default using the `set` command, as in

```
=> set refer
```

All subsequent requests for displaying records will be in the newly requested format.

If you would like to display the current results in an alternative format without changing the default format, add the format name at the end of the list request. The following command will display the current results in `long` format, for instance, but will continue to use the default format for any subsequent requests.

```
=> list long
```

2.11 Saving to a file (`save`)

If it is desired to save the results of a search on a disk, the `save` command will prompt you for a file name, and then CIScd will copy the contents of the main search results to the file with that name. If the file name contains spaces or special characters, or if it specifies a full path name, then the file name should be enclosed in double quotation marks (`"`). This makes it possible to save results to a floppy disk, for instance. Here are two examples:

```
=> save "ridge regression biblio"
=> save "a:\papers\searches\ridge.cis"
```

The save file is written using the current default format. You can specify a different format for saving the file by adding the the format name to the end of the `save` request, as in this example:

```
=> save best.cis refer
```

2.12 Loading a saved file (`retrieve`)

It may be particularly useful to save the results of a long search, for further refinement (or perhaps for using as a fresh starting point later on). To do this, save the current search in raw format. Any file that is saved in raw format can be loaded into the main search results by using the `retrieve` command. Thus, the contents of the main results will be the same both before and after the following sequence of commands is given.

```
=> save best.raw raw
=> .....
=> retrieve best.raw
```

2.13 More (`more`)

Records that scroll off the top of the screen cannot be looked at (on MS-DOS machines) without doing another search. For this reason, the `more` command can be used to cause screen output to stop after roughly every 24 lines of output. The next screen of output is produced by typing the Enter or Return keys on the keyboard. This paging feature can be toggled on and off by repeated uses of the `more` command.

Alternatively, one can use the following commands to control this feature.

```
=> set more            (enables paging)
=> set nomore          (disables paging)
=> more 60             (stops after every 60 lines or so)
```

2.14 Screen display (`list`, `display`)

The `list` and `display` commands show the results of the current main search at the screen. These names are synonyms. The output format can be selected as described above under "Output format".

2.15 Showing the next *k* records (`next`)

A convenient way of looking at a set of search results is to display from two to four records at a time, depending on the display format currently in effect. The command

```
=> next 4
```

will cause the next four records to be displayed each time the command is issued. If the last request issued was a `next` command (either with or without a following integer) an empty command line will be treated as equivalent to the `next` command. Thus, setting `next 3` followed by an empty line will display the next three items in the current search. Every time the Return key is pressed another three records will be displayed. Any command other than `next` or Return will cause the Return key to revert to its usual function.

2.16 Manipulating the search-result stack

The `stack` command gives a somewhat cryptic status report of the contents of the stack of search results. The `push` command and `pop` command move the stack contents by one position.

2.17 Displaying the contents of a disk file (`type`)

The `type` command will copy a named disk file to the screen.

2.18 Displaying settings

The current values of all options that can be `set` are displayed with the `show` command.

2.19 Version

The `version` command displays both the current version of the software, and a brief description of the database being examined (as told by the configuration file).

2.20 Comment (`#`)

It is sometimes useful to put comments between commands. Any command line beginning with "`#`" is merely printed and ignored. Blank lines are also ignored (except immediately after a `next` command).

2.21 Looking up an abbreviation (`whatis`)

Not every abbreviation used in the *CIS/ED* is self-explanatory. The `whatis` command will look up the abbreviation in the abbreviation mas-

ter list (provided that CIScd has been configured with an abbreviation file):

```
=> whatis CmpBiomd
   Abbreviation 'CmpBiomd' corresponds to
   Computer Methods and Programs in Biomedicine
```

The command may be abbreviated to its first four letters.

2.22 Using CIScd with a printer (DOS/WINDOWS only)

Portions of a search session can be printed to an attached printer using the built-in capabilities of your computer.

To print a copy of the current page displayed on your screen, press the CONTROL and PRINT SCREEN keys simultaneously.

If you would like to capture a transcript of your session (or portions of it) to the printer, typing CONTROL-SHIFT-PRINT SCREEN will send a copy of each subsequent line typed or displayed on the screen to the printer. This will continue until CONTROL-SHIFT-PRINT SCREEN is pressed again to disable the feature.

Many printers do not begin printing a physical page until the page's entire contents have been sent to it by the program (Hewlett-Packard inkjet printers are an example). In this case, pressing the reset button on the printer once you have finished sending it the material that you wish to print will initiate actual printing.

2.23 Restarting, or changing configurations

Although this is rarely needed or desirable, it is possible to start the program again from the beginning, optionally reading a specified configuration file. The commands (in these two variants) are:

```
=> config             # to restart using the default configuration
=> config alt.cfg     # to restart using alt.cfg
```

2.24 Quit (`quit`)

This command will exit the program. On the Macintosh, the window from the search session will remain active, so that it can be printed and/or saved to a file. Cut and paste are also active, and after the `quit` command is

issued, the contents of the window can be edited or formatted (to delete dead ends or errors in the search, for instance). To completely quit on the Macintosh, the **Quit** option must be selected from the **File** menu (or COMMAND-Q must be typed from the keyboard).

CIScd creates temporary files to hold the results of searches and refinement requests. These temporary files are not deleted on quitting the program unless you specifically ask for them to be removed. (This makes it possible to exit from the program, and then to start the program again later, taking up exactly where you left a search project. If you are certain that you will not need the current intermediate results, you can give the command

```
=> quit clear
```

which will delete the work files. If you do not specify `clear`, you will be asked whether you wish to delete the files. In this case, the files will be deleted only if you explicitly request them to be.

2.25 Modifying the `cis.cfg` when Files are on Hard Disk

Documentation for this feature is currently available from the *CIS/ED* website at `http://www.stat.uchicago.edu/~cis`.

2.26 Fine Print

The software and the documentation contained in this section are copyright: Copyright © 1994–7 Ronald A. Thisted. All Rights Reserved. Portions of the program for the search command are Copyright © 1993 by Douglas M. Bates.

No warranty of any sort is provided with this software. Authorized users of the *Current Index to Statistics* Extended Database (*CIS/ED*) are granted a nonexclusive royalty-free license to use this software in conjunction with the *CIS/ED*.

CIScd is provided to you at no cost. You may make as many copies of CIScd as you (or individuals in your organization, if you have an institutional *CIS* Database license) can personally use. You may not sell or redistribute the program in any other manner, except to recover the actual cost of the diskettes you may use to provide the program to others.

Part III
User Reference

3 Overview

The Current Index to Statistics is an organization sponsored jointly by the American Statistical Association (ASA) and the Institute of Mathematical Statistics (IMS). Under the direction of a Management Committee, the organization produces an annual printed index to the statistical literature appearing in the previous year, and a cumulative database issued in electronic form. The 1997 edition of the *CIS Extended Database* (*CIS/ED*) contains approximately 28 Mb of bibliographic information, including entries from Volumes 1–22 of the *Current Index to Statistics* (covering the statistical literature published in the years 1975–1996). This edition of the *CIS/ED* also includes many pre-1975 journal articles not originally contained in the printed *Current Index to Statistics* volumes (called "Gap" records below), more complete coverage from the early years of the printed volume, and some electronic abstracts.

3.1 Organization and Contents

Information from each year is organized into a single file. For example, the file for 1987 is called `CIS87.V97`. The items in the files for each year are the ones with that year as their nominal publication date, even if some of the items did not appear in the printed CIS volumes until later years. In this edition, the 1996 files are shorter than the files for other recent years because some "1996" publications appeared too late for inclusion in this edition. Files containing records from publications prior to 1965 are contained in a file named `CIS0064.V97`.

In addition to the bibliographic files organized by years, each copy of *CIS/ED* contains

- Source abbreviation files,
- Copies of the license agreement,
- A computer-readable version of this *User Guide,*
- A computer-readable version of the *CIS/ED Technical Reference Manual*,

- Abstracts for some referenced articles,
- Text of book reviews from ASA journals published after 1995,
- Contributed software, and
- Approximately 17 Mb of pre-compiled index files for use with search software,
- Other documentation.

These items are in the subdirectories corresponding to the supported computer platforms: Macintosh, MS-DOS, and Unix™. They are located in directories `ABBREV`, `LICENSE`, `USRGUIDE`, `CONTRIB`, `INDEX`, and `DOCS`, respectively.

CIS/ED consists of several "flat" databases. The main database consists of individual *records* separated by end-of-line characters. There are separate versions for Macintosh, MS-DOS and Unix™ computers, with the line separators appropriate to the respective computers. Each record pertains to one indexed paper, book, database, or computer program. There are about 200,000 records altogether. Each record is divided by eight `#`-characters into seven *fields* of variable length (except for the first), some of which may be empty. Multiple *entries* (or *subfields*) within fields are separated by semicolons. The contents of these fields are described in detail in Chapter 4. To use the database well it is important to understand how the information in each field is organized.

Other databases included in the distribution, also with a "flat" structure, are data files whose records contain information about

- Abbreviations used for publishers and edited volumes
- Abstracts of selected articles appearing in the main database
- Indexes to the main database

These flat files can, in principle, be linked using keys in common across databases, so that they can be configured into a relational database. Contributed search software may free the user from having to know anything about the specific formats for the various databases, by automatically retrieving all relevant information relating to a record and exhibiting it in a standard bibliographic format, such as Refer or BibTeX.

Descriptions of the abbreviation files, abstracts, and special indexes are contained in the *CIS/ED Technical Reference Manual*.

3.2 Search Software.

Search software contributed by several different authors for Unix™, DOS, Windows, and Macintosh computers are included on the CD-ROM. Although the *Current Index to Statistics* Management Committee does not itself support, license, recommend or warrant search software, it does encourage software development, makes space available on the CD-ROM for software contributed from a variety of sources, and provides pointers to software on its Web page.

The CD-ROM contains platform-specific directories that contain contributed software for searching the database. All of the search software distributed on the CD-ROM (or a more current version of that software!) is also available on the *CIS* Web site (`http://www.stat.uchicago.edu/~cis`) and on `Statlib`. *Please consult the CIS Web page for updates, revisions, and the latest version of search programs for CIS/ED.*

The directories in which software can be found are `DOSSFT` for DOS-specific programs, `WINSFT` for Windows-specific programs, and `UNIXSFT` for programs intended for the Unix™ environment. Macintosh users will find contributed software in the "Current Index Software" folder. Pointers to other sources of search software are contained in the directory `OTHSFT`.

All of the software for DOS, Windows, Macintosh, and Unix™ computers contained with the *CIS/ED* has been contributed by users, in the hope that some *CIS/ED* users will find it helpful. Users who develop or can recommend helpful software are encouraged to contact the editor of the database. All contributed software (freeware or shareware) submitted to and approved by the Editor will be included in the 1998 CD-ROM distribution. Consult the *CIS* Web page (`http://www.stat.uchicago.edu/~cis`) for information for software developers, including schedules for submission.

Disclaimer. Any software distributed with *CIS/ED* is provided as a service only. The *Current Index to Statistics* does not support or recommend any search software. The software supplied on the CD-ROM is copyright by the individual authors. As a condition of use, users of the software must comply with any licensing or other restrictions imposed by the software's authors.

3.3 User Guide

The *User Guide* can be found on the CD-ROM, in the directory `USRGUIDE`. There are three versions: LaTeX (file `USRGUIDE.DVI`), Adobe Portable

Document Format (file `USRGUIDE.PDF`), and ASCII (file `USRGUIDE.TXT`). The LaTeX and Adobe Portable Document Format versions are identical in content to the bound volume. Adobe Portable Document Format versions of documentation require Adobe *Acrobat* version 3.0 or higher. This free document reader can be obtained from Adobe's web site at `http://www.adobe.com/`. The ASCII version is unpaginated; it can be used to browse the *User Guide* on-line, say with a text editor. The ASCII version is necessarily only a rough approximation to the bound volume, which should be consulted to settle any ambiguities.

Any of these versions may be printed and duplicated.

3.4 Technical Reference Manual

Some information, previously contained in the *User Guide*, has been moved to a new *CIS/ED Technical Reference Manual*. This document can be downloaded from `http://www.stat.uchicago.edu/~cis` in Adobe Portable Document Format. It is also provided on the CD-ROM.

The *Manual* contains technical specifications and technical information on the CD-ROM format, the detailed specifications of formats used for fields 1 and 6 of the database records, specifications for the abbreviation format files, specifications for abstracts, LaTeX styles for printing *CIS* abstracts, and information about Russian and other names transliterated from Cyrillic characters.

3.5 Changes in Future Editions

Your suggestions for improving *CIS/ED* are welcome. Also, the correction of errors is an important continuing effort in which user cooperation is essential.

Several important changes will be implemented in future editions of *CIS/ED*. We solicit comments from users about these items.

3.5.1 Theses and Technical Reports

Considerable interest has been expressed in title, author, and key word information for theses and technical reports. In order for this effort to be financially feasible, for university departments and research groups must submit records electronically to the editors. Current information, submission standards, and instructions can be found at our Web site, `http://www.stat.uchicago.edu/~cis`.

If any thesis or technical report records are submitted, they will be included on the CD-ROM in separate files. Records with such information would not be intermingled with records on journal articles, books, etc.

3.5.2 Reviews

Beginning with the 1998 edition, book reviews will be distinguished from journal articles in Field 1 using a new code for reviews in column 28 of Field 1.

3.5.3 Non-English Titles

In future editions, we hope to include the title of articles *in the original published language*, where possible, as well as an English translation. The keyword field(s) would contain English-language keywords that may overlap words and phrases that would appear in an English translation of the title.

4 The Database Records

This chapter describes the information that may be contained in each of the seven fields of a record. The first four fields are used in creating the printed version of the *Current Index*. The remaining fields are included in *CIS/ED* to make searches easier and for administrative purposes.

Three illustrative records taken from `CIS87.V97` are shown below, just as they appear in the database (except for the line-breaks needed to print them here):

```
#87Academic:NY:Lnd     442p B#Probabilistic methods in
mathematical physics#It\^o, K.;Ikeda, N. (Ed)##Ito,
K.##PV15#

#87AnlsStat 15    1155-1173 J#Convergence rates for the
bootstrapped product-limit process#Horv\'ath,
Lajos;Yandell, Brian S.#Lorenz curve;Censoring;Product
limit estimator#Horvath, Lajos##PV13;IMSC#

#87AplMaCmp 24      47-  63 J#Regression estimates of
inputs to an M(t)/G/$\infty$ service system#Patterson,
Richard L.#Queueing###PV13;94m#
```

4.1 Field 1: Location

This field contains information on the year of publication, source, and location or size of the item indexed. The order of records within each file is determined by a case-insensitive alpha sort on this field. (That is, all letters are temporarily changed to upper case before sorting.) The file `CIS87.V97` contains records showing a 1987 publication date.

Elements within Field 1 appear in fixed column locations. Several different formats are used, depending on the type of publication being referenced. Field 1 is always 28 columns in length. Columns 1–2 always contain a two-digit year of publication, and column 28 always contains an indication of the publication type, using the codes in Table 1. If an abstract or other short description or summary is available for the record within *CIS/ED*, the code in column 28 is replaced by its lowercase equivalent.

The *CIS/ED Technical Reference Manual* describes the formats in detail. This document can be obtained from the CD-ROM and from

http://www.stat.uchicago.edu/~cis. A brief synopsis of the contents of Field 1 follows.

Table 1: Publication Types

B	Book
J	Article in journal or other periodical (not electronic)
P	Article in proceedings or edited book
C	Computer-readable databases, software, etc
E	Article or entry in electronic journal or periodical
Z	Administrative record (no bibliographic information)
D	Dissertations (not used in 1997)
T	Technical Reports (not used in 1997)

B Columns 3–18 contain a publisher abbreviation, 19–20 spaces, 21–25 number of pages, 26 the letter p, 27 an optional character used, if necessary, for uniqueness.

J Columns 3–10 Journal abbreviation, 11–13 volume, 14–16 issue number, 17–21 beginning page, 22 hyphen, 23–26 ending page, 27 an optional character used for special circumstances.

P Columns 3–12 Abbreviated title of edited book, 13–16 "volume" number, if any, 17–21 beginning page, 22 hyphen, 23–26 ending page, 27 optional character.

C Columns 3–18 Publisher, producer, or distributor abbreviation, 19–20 spaces, 21–24 size of file, 25–26 kb, mb, or gb to denote size units, 27 optional character.

E Columns 3–10 Journal abbreviation, 11–13 volume number, 14–16 issue number, 17–25 publication-specific item-locating information, 26 code indicating format used in columns 17–25, 27 Optional character.

Z Columns 1–27 are reserved.

4.2 Field 2: Title

This field shows the title of the indexed item as it appeared in the source. When non-English titles are translated into English, the name of the original language is given in parentheses. In some records, full titles are given

in the original language in parentheses after the English translation. Parenthetical comments may also show the location of discussions, correction notes, addenda, or acknowledgements of priority. Mathematical notation is represented in TEX codes (Knuth, Donald E. (1986). *The TEXbook.* Addison-Wesley: Reading, Mass).

4.3 Field 3: Authors

Authors' names are given in the spelling shown in the source, except to correct confirmed typographical errors in the source. Multiple authors are separated from one another by semicolons. Different formats are used for corporate or institutional authors and for individuals.

4.3.1 Corporate Authors

Corporate or institutional authors are uniquely identified by a leading blank space in their names. In contrast to the syntax for individual names, any commas in the names are used merely for proper punctuation. Here are some examples. Note that each name begins with a single space, which we have emphasized with the "␣" symbol.

```
␣Akademie-Verlag
␣ASA Subsection on Teaching Stat in Health Sciences
␣U.S. Bureau of the Census
␣Panel on Discriminant Analysis, Classification,
                                          and Clustering
```

Names may be followed by a ROLE ABBREVIATION in parentheses, as discussed in Section 4.3.3.

4.3.2 Individual Authors

Authors' names are given in inverted order. The syntax for an individual author's name is:

- SURNAME (or family name),
- If necessary, a comma followed by FIRST NAMES and/or initials,
- If necessary, another comma followed by a SUFFIX (such as Jr, III, Esq, etc),
- If necessary, a ROLE ABBREVIATION in parentheses.

Thus, a person's name may have zero, one, or two commas. Initials in names are always followed by exactly one period. Initials are separated from subsequent text by a single blank space following the period (unless the next character is a comma, semicolon, or `#`-character). Here are some examples of names (without role abbreviations).

```
Editors
Bhagawandas
Chan Fu-Chen
Sanathanan, Lilly
Blattberg, Robert C.
Brown, L. D.
De\'ak, Istvan
Erd\H{o}s, Paul
Le Breton, A.
van der Horst, Charles M.
Mu\~noz P\'erez, J.
Garcia-Carrasco, Pilar
Harrell, Frank E., Jr
Pickands, J., III
```

Records originating with *CIS* Volume 14 (1988) and later (and all Gap records) have TEX codes for accent marks on authors' names, and some earlier entries have also been edited to include correct accent marks.

4.3.3 Role abbreviations

Parenthetical roles — `(Auth)`, `(Ed)`, `(Rev)`, `(Transl)`, for author, editor, reviewer, and translator respectively — follow all other components of a name. A few other roles are also used; a full list can be obtained from the editors. Role entries have been standardized, so that the role abbreviations are now consistent across all records. In addition, the order in which roles are listed is standardized, so that authors precede editors, who precede reviewers. Translators appear last.

The role `(Auth)` is used to indicate the author(s) of a reviewed work, to distinguish them from the author(s) of the review itself. The latter are indicated by the `(Rev)` notation.

For conciseness of presentation, the role is included in the database entry only after the last person who plays each role. For example, a book

by Efron and Tibshirani reviewed by Jones and Smith would have the following entry in Field 2:

```
Efron, B.;Tibshirani, R. (Auth);Jones, J.;Smith, S. (Rev)
```

4.4 Field 4: Key words and phrases

These may be edited versions of key words published in the source or they may have been prepared by a *CIS* editor. Beginning in 1997, Field 4 for some journals also contains authors' key words which were not selected by the *CIS* editors. These words and phrases are enclosed in [square brackets], and they follow the editors' key words. Some articles appear in a named department, such as the "Teacher's Corner" for the *American Statistician.* These department names may also be included in Field 4; if so, they are contained in square brackets, and they preced the editors' key words.

The tendency (but far from a fixed rule) has been to use singular and non-possessive forms (e.g., Chebyshev polynomial rather than Chebyshev's polynomials). Words and phrases appearing in the title are not repeated in Field 4.

4.5 Field 5: Searchable author information

Since 1988, TEX codes for accents on authors' names have been included in Field 3. These codes should in most cases produce the appropriately accented version of the name for printing purposes.

Unfortunately, authors' names can be accented, transliterated, and even spelled in very different ways. These variations in presentation make comprehensive searching difficult. For this reason, Field 5 contains the following kinds of author entries to make searches more likely to be successful:

- Names with accent marks and other symbols removed. These are called STRIPPED names, and consist of the original name from Field 3 with all TEX codes and apostrophes removed. For example:

In Field 3	*In Field 5*
`Muñoz`	`Munoz`
`O'Brien`	`OBrien`

- Names RESPELLED using standard roman transliteration. For example:

In Field 3	*In Field 5*
Chadœuf	Chadoeuf
Müller	Mueller
Bjørk	Bjoerk
Åke	Aake

- Standardized TRANSLITERATION. Many names originally spelled in Cyrillic or similar alphabets have been repeated in Field 5 with standardized transliterations. For example

In Field 3	*In Field 5*
Hasminski	Khasminskii
Has'minsky	Khasminskii
Kha'sminskiĭ	Khasminskii

  See the *CIS/ED Technical Reference Manual* for more information.
- Suspected CORRECTIONS. In a few instances where we suspect that an author's name may have been misspelled in Field 3, we have put a suggested correction into Field 5 (usually followed by a question mark in parentheses).

It is important to note that the entries in Field 5 are provided to enhance searchability, and are often not displayed by search software. Most entries in Field 5 are automatically generated by algorithm to enhance both human- and computer-generated search terms.

4.6 Field 6: Additional Searchable Information

Key words/phrases in addition to those in Field 4 are placed in Field 6. These include unaccented (stripped) versions of words and phrases from Field 4, noun phrases giving unaccented information from Field 2 (title), phrases with Americanized spelling of words or phrases that are likely to be the objects of searches, etc.

4.7 Field 7: Administrative codes

Field 7 is used by the *CIS/ED* editorial staff. Search software ordinarily suppresses Field 7.

5 Hints for Searching CIS/ED

Some general principles

Consider starting with a search that will retrieve too much and finding ways to trim the result to what you want. This sometimes requires less imagination and skill than retrieving too little at first and then trying to figure out how to expand the inadequate result. Note that this principle can be implemented cheaply with a locally-installed database such as this one, but it might be quite costly with an on-line system where you are paying for each search by the second or for each item retrieved.

In planning a search, you may find it helpful to first use a text editor to browse the subject, keyword, or author indexes (see Chapter 7). This will give an idea of the number of variants of a term, as well as the relative frequency of each variant in the database.

If you are using contributed search software, you should become familiar with the search strategies that it recommends.

Beware of hyphens versus spaces

Hyphens and spaces are used quite inconsistently and often even incorrectly in the titles of journal articles. These inconsistencies are reflected in the *Current Index to Statistics*, which records titles as they appear. As a result you cannot predict, for example, whether the record you seek will have `log linear model`, `log-linear model`, or even `log linear-model`. In such a case it is best to search for records which contain both `log` and `linear`.

If your search software uses an inverted index, it can retrieve only the precise "words" that its authors have placed into the inverted index. For example with such software you may need to retrieve the union of `loglinear`, and `log` intersect `linear`.

Here is an example of such a search using the CIScd program:

```
search log linear
search loglinear
merge
```

Case sensitivity

When the search program is capable of doing so, it is usually best to do case-sensitive searches for author names, and case-insensitive searches for key words and title words. When doing a case-sensitive search, note that the first word of a title is always capitalized in *CIS/ED*. Subsequent words are capitalized only if they are, or are derived from, proper nouns, e.g., `Monte Carlo`.

The inverted indexes supplied with *CIS/ED* are not case-sensitive.

Greek letters and mathematical notation

Do not attempt to use TEX codes in search terms. Because the TEX code for a mathematical notation is seldom unique, and because some journals spell out Greek letters (phi-mixing) while others use symbolic notation (ϕ-mixing or φ-mixing), the database contains in Field 6 a version of such words stripped of TEX notation. The file `F6.V97` can be scanned to confirm the standardized version used in Field 6. Some examples are `phi-mixing`, `n by n Latin square`, `Cp-statistic`, and `2 by 2 table`.

Author's names

The names of people are listed in Field 3 in inverted form, with the family name(s) appearing before the given name(s) (see Section 4.3.2). Thus a search for `Stein, C` will retrieve `Stein, C.` and `Stein, Charles` and `Stein, Charles M.`, but not `Stein, Michael` or `Stein, William` or `Steinberg`, etc.

Multi-part family names and firm names

Some names are particularly likely to cause difficulty. Consider, for example, `Willem van der Velt`, `Carlos Mendez Gonzales`, `Park Dong Ho`, and `SPSS, Inc.` Some sources may have `van der Velt`, but others may have `vander Velt` or `Van Der Velt`, and these variations would be reflected in Field 3 for the corresponding records. Similarly, it is possible to find entries under `Mendez-Gonzalez`, `Mendez y Gonzalez`, `Dong Ho Park`, and `SPSS Inc.` (note the missing comma between the two words). Even if you know which form is "correct" — or which form the author prefers — you cannot be sure that either the original journal or *CIS/ED*

has it right. It is best to scan the author index before doing any comprehensive searches for an author whose name might have several variants (see Section 7.2). It may be helpful to search on a single word in the name (unless that is likely to yield too many unwanted records) or to use intersections; for firm names, don't include `Inc.`, `Ltd.`, etc. (with or without a preceding comma) in a search string, since there is little consistency in usage.

Correction notes, addenda, etc.

Authors may submit a correction note, addendum, or acknowledgment of priority months or years after the original paper appeared. The location of a correction note is indicated in parentheses after the title of the original paper to which it refers.

Discussions

Some papers are immediately followed by comments (and often the authors' replies). Such remarks can usually be found by searching on a phrase in the title of the main paper. Many entries indicate the presence of such discussion in parentheses after the title of the main paper, especially if the discussion appears immediately after the paper. (Look for notations beginning with `Disc:`, `Com:`, `C/R:`, or `Pkg:`. Except for `Pkg:`, which denotes a package of related papers and discussion, page numbers in these notations are not included in the span of pages for the main paper.)

Some journals may publish comments (often as letters to the editor) several years after the original paper. Often, especially for more recently created records, a parenthetical notation (`Disc:`, `Com:`, `C/R:`) after the title points ahead to such discussion. However, if you know that a particular journal has a policy of encouraging delayed discussion, and if you have retrieved a paper by key words rather than by words in its title, you may want to search by title for detached discussions.

The distinction between "comments" and a correction or addendum is authorship; comments are contributed by persons other than the author(s) of the referent paper. *CIS/ED* uses the words Comments and Reply even if the journal calls these remarks Discussion and Rejoinder or gives them separate titles of their own.

Accent marks

Eliminate all accent marks in searches. Similarly, for transliterated names omit any apostrophes (that is, `Silvestrov` not `Sil'vestrov`). Fields 5 and 6 have been created to make this easy approach work well.

Exceptions. For German umlauts it is best to use the spelling with `ae`, `oe`, or `ue` so that you will retrieve both pre-1988[1] and later records, but be aware that if the source used neither an umlaut nor the spelling with the extra e, then the briefer spelling alone will appear in the *CIS/ED*. Although in the 1997 edition, we have made an attempt to systematize usage, the pre-1988 practice for non-German umlauts/diereses and certain Scandinavian letters does not appear to have been consistent, so for these it is still safest to search on both forms (Examples: `Noel` and `Noeel`, `Bjorn` and `Bjoern`, `Ake` and `Aake`.) Also, journals sometimes print "Hungarian" umlauts as "German" umlauts, so you may see `Csoergoe` as well as `Csoergo` (for Csörgő).

Transliterated Cyrillic Names

Standardized spelling of many Russian and other names which have been transliterated from Cyrillic texts, such as `Chebyshev`, `Gikhman`, `Markov`, `Prokhorov`, and `Skorokhod`, have been used in Field 5. A list of the names for which standardized spelling has been introduced is contained in the *CIS/ED Technical Reference Manual*.

Miscellaneous

For important words that can be spelled either with `-or` or with `-our`, use the U.S. spelling. In other cases where U.S. and British spelling differs, it is best to try both search strings. The abbreviations `U.S.`, `U.K.`, `U.S.S.R.`, `U.N.`, and `W.H.O.` do not contain spaces in Fields 2, 3, and 4; they should always contain periods. For the convenience of those using software based on inverted indexes (which may treat periods as word breaks even if not followed by a space), instances that contain periods have been repeated in Field 5 or 6 as `USA`, `UK`, `USSR`, `UN`, and `WHO`.

[1] Before 1988, the printed volume of *Current Index to Statistics* did not use TEX codes for umlauts, but transliterated them in the customary fashion.

6 Source Abbreviations

The journal and publisher abbreviations used in Field 1 (called "keys" below) are those used in the annual printed *Current Index to Statistics* volumes. Many keys are easy for statisticians to decode, e.g., `AnlsStat` or `JASA`, however others may be mysterious. The database contains several files that give expanded forms of the keys. These files are located in the `ABBREV` directory of the CD-ROM, and are described in detail in the *CIS/ED Technical Reference Manual*. Here is an overview of the files.

`ABBRMST.V97` This has the most complete information about each key. However it is not well-suited to human browsing because of the extensive use of cross-references to other keys.

`ABBR.V97` This is perhaps the most useful file. It gives the full name associated with each key, e.g., `The Annals of Statistics` for `AnlsStat`. The publishers of journals and edited books are identified, as are the editors of most edited volumes. `ABBR.V97` gives information similar to that in Parts 4 and 5 of each annual printed *Current Index to Statistics* volume, but it is comprehensive for the entire span of *CIS/ED*.

`ABB.V97` This file is similar to `ABBR.V97`, but uses abbreviated forms of names, e.g., `Ann. Statist.` for `AnlsStat`. These short forms are similar in style to those used in *Mathematical Reviews*, published by the American Mathematical Society. There are some differences, because the contractions used in *CIS/ED* were chosen to be consistent across a much broader range of statistical publications than is found in *Mathematical Reviews*.

`AB.V97` This file is similar to `ABB.V97`. It is easier for humans to look at, but harder for computers to parse.

`ABAMS.V97` This file is related to `ABB.V97` and `AB.V97`. For those publications and publishers that are represented both in *CIS/ED* and *Mathematical Reviews*, this file gives the short form (e.g., `Ann. Statist.`) of the name used by *Mathematical Reviews*.

`ABBRAMS.V97` This file is similar to `ABAMS.V97`, but gives the long form (e.g., `The Annals of Statistics`) of the name used by *Mathematical Reviews*.

7 Special Indexes

The 1997 edition of *CIS/ED* contains a Subject Index, a Keyword Index, a list of additional searchable terms, an Author Index, two sets of Inverted Indexes, and a Thesaurus of related phrases. The indexes and Thesaurus are located in the platform-specific directories of the CD-ROM, in the subdirectories `INDEX` and `INDEX2`.

7.1 Subject Index

The subject index (`SUBJECT.V97`) is an ASCII file, approximately 0.8 Mb in length, tabulating the number of records in which various "words" occur somewhere in Fields 2 (titles), 4 (key words/phrases), or 6 (additional searchable information).

The subject index can be used to browse the contents of the database, to check for alternative spellings of words, and to locate the occurrence of names included as parts of titles. For instance, the subject index can be used to identify occurrences of `Stein-James`, `Stein`, and `James-Stein`, all of which potentially refer to the 1955 shrinkage estimator of Charles Stein.

Subject Index Format

The format for the Subject Index is intended to make browsing with a word-processor or screen-oriented text editor relatively easy. Because many words and phrases in titles and keywords contain TEX codes, the non-mathematical entries in the Subject Index are sorted by a TEX-free version of the name (the "sort-word") obtained by stripping all TEX codes and special characters from the word or phrase (the "actual-word"). Records in the subject index have three fields: count, sort-word, and actual-word. The last field is empty if the actual-word is the same as the sort-word, as is commonly the case.

The count field is of fixed length (five characters), the two remaining fields are of variable length, and the fields are separated by the TAB character (hexadecimal code `09`). Each entry is terminated with a newline character.

The index should be searched in a case-insensitive manner. However, for the sake of readability, some words in the index appear in mixed case. This is true of most acronyms (for example, `ARMA`, `SAS` and `EM`), and most

proper names (for example, `Markov`, `Fisher`, and `Egdeworth`). Occasionally a word appears in both upper and lower case (for example, `BLUE`, an acronym, and `blue`, presumably the color). Here are some examples, taken from the 1996 Edition of the *CIS/ED*:

```
   18 $R \times C$
   18 $r \times c$
  380 $t$
  124 $t$-distribution
  136 $t$-test
   16 $t$-tests
 2745 (French)
19777 analysis
    8 analysis-of-variance
  684 ARMA
   15 ARMA$(p,q)$
   15 ARMA(p,q)
   58 BLUE
   10 blue
    1 BLUES
    6 BLUEs
   37 French
  366 goodness
  942 goodness-of-fit
    1 goodness-of-fit-tests
   10 iid
    2 iid i.i.d
   73 iid i.i.d.
  142 M/G/1
  249 meta-analysis
   60 Scheffe Scheff\'e
```

Generating the Subject Index

The computer algorithm used to generate indexed words is too complicated to describe in detail here. Roughly speaking, the process involves:

1. extracting textual fragments containing TEX math mode material (for example, `$r \times c$`),
2. breaking the remaining text on blanks and semicolons,

3. stripping most leading and trailing parentheses and quote marks (including apostrophes and single quotes), and

4. discarding some items not likely to be used in searches (such as most numbers, including page ranges and volume numbers)

As a result, the subject index contains:

- all ordinary words, e.g., `analysis`, `estimation`, and `French`;
- all hyphenated phrases, e.g., `goodness-of-fit`, `chi-squared`, and `P-P`;
- all textual fragments containing TEX math mode, e.g., `$t$` and `$M$-estimator`;
- words with imbedded non-alphanumeric characters or TEX accents, e.g., `M/G/1`, `SAS/GRAPH`, `AR(p)`, and `Scheff\'e`;
- the language of publication (if different from English), enclosed in parentheses, e.g., `(French)` and `(German)`; and
- a few phrases involving numbers (e.g., `0-1` and `2-dimensional`).

Some words in the index occur in two forms — the way the word actually appears in the database (for instance, `Scheff\'e`, or `i.i.d.`), and the version of the word used for purposes of sorting (e.g., `Scheffe`, or `iid`).

7.2 Author Index

The author index (`AUTHOR.V97`), approximately 2.5 Mb in length, is designed to fill a similar role to the subject index. For the author index, entries in the author field (Field 3) are extracted and sorted by their "stripped" versions, that is, the name with all TEX codes and special characters removed. Each author is tabulated separately in each of several possible *roles*, encoded by a single character as shown in the table below. This makes it possible to distinguish indexed items written by an individual (such as articles, books, or book reviews) from indexed items written by someone else about a work for which the individual was author or editor (such as a reviewed book or computer program). The role codes are given in the table below:

```
A = Author of indexed item
a = Author of reviewed work
D = Discussant
E = Editor
R = Reviewer
T = Translator
```

Author Index Format

The format of the Author Index is intended to facilitate browsing the file with a text editor or word processing program. Each entry is encoded in ASCII and consists of two fields, separated by a TAB character (hexadecimal `09`). The first field is exactly six characters in length, and contains a count (four characters), a single blank space (1 character), and the role code (1 character). The second field consists of the stripped version of the name being indexed, and is followed in parentheses by the TEX-encoded version from which it came, if appropriate. Each entry is terminated by a newline character.

Examples

1. The following example illustrates several uses for the Author Index. To investigate the works of Murray Aitkin, one might start by perusing the following section of the Author tabulation. (The examples in this section are from the 1996 Edition of the *CIS/ED*.)

```
   1 A Aitken, M.
   1 a Aitken, Murray
   6 A Aitkin, M.
   6 a Aitkin, M.
   3 R Aitkin, M.
   5 A Aitkin, M. A.
   1 a Aitkin, Murray
  33 A Aitkin, Murray
   4 R Aitkin, Murray
   1 A Aitkin, Murray A.
   1 R Aitkin, Murray A.
   1 A Aitkin, Murry A.
```

It is immediately apparent that some entries, which *probably* represent works by Aitkin, are listed under "Aitken" or "Murry." It is not obvious whether this state of affairs is due to a precise transcription (by *CIS*) of an error appearing in the original journal, due to a typographical error introduced by *CIS*, or due to a second (or third!) individual having a remarkably similar name to the person in whom we are interested. A careful scholar will investigate these entries by using standard CIS search methods to identify the records and then by checking with original sources for additional information.

2. The following excerpt from the index illustrates the treatment of names which have been transliterated from the Cyrillic alphabet using different transliteration methods.

```
1 A Aleshkyavichene, A.
1 A Aleshkyavichene, A. (Ale\v{s}kevi\v{c}iene, A.)
1 A Aleshkyavichene, A. D.
7 A Aleshkyavichene, A. K.
1 A Aleshkyavichene, A. K. (Aleshkevichiene, A. K.)
2 A Aleshkyavichene, A. K. (Aleshkiavichene, A. K.)
2 A Aleshkyavichene, A. K. (Aleskeviciene, A. K.)
3 A Aleshkyavichene, A. K. (Aleskjavicene, A. K.)
1 A Aleshkyavichene, A. K. (Alyeshkyavichene, A. K.)
1 A Aleshkyavichene, A. K. (Alyeshkyavichenye, A. K.)
```

   Here, the entry in parentheses is the actual form of the name, as found in Field 3. The main entry is a "canonical" transliteration based on the American Mathematical Society system; no TEX codes are used. Since the canonical transliteration is also inserted into Field 5, a search for "`Aleshkyavichene, A`" would recover *all* of the records indicated above.

3. The final example illustrates the manner in which TEX-ed names are converted to a common stripped key.

```
1 A Gonzalez, W. (Gonz\'alez, W.)                          *
1 A Gonzalez, Wenceslao                                    *
1 A Gonzalez-Andujar, J. L.
1 A Gonzalez-Andujar, J. L. (Gonz\'alez-Andujar, J. L.)
2 A Gonzalez-Farias, Graciela
```

```
1 A Gonzalez-Garcia, C. (Gonz\'alez-Garc\'{\i}a, C.)
1 A Gonzalez-Guzman, Jorge (Gonz\'alez-Guzm\'an, Jorge)
1 A Gonzalez-Manteiga, W.                              *
5 A Gonzalez-Manteiga, W. (Gonz\'alez-Manteiga, W.)    *
1 A Gonzalez-Manteiga, Wenceslao                       *
```

Here, the parenthesized version denotes the author's name as found in Field 3, while the main entry indicates the stripped form placed into Field 5. By visually scanning the Author Index in the neighborhood of an author of interest, variants of the author's name can sometimes be found. In this case, it is likely that the five entries marked with an asterisk above were all written by the same author, but no single search strategy would easily find both forms of the name. Forearmed with this information, an investigator can search in the *CIS/ED* under both variants.

7.3 Index of Key Words and Phrases

The keyword index (`KEYWD.V97`), approximately 0.5 Mb in length, is a tabulation of the semicolon-separated key words/phrases in Field 4 of the database. The two fields, "count" and "term", are separated by a blank; the count is right justified in a field of width 4. The count does *not* include occurrences of the term in the title field (Field 2) of database records.

For example, the following entries from `KEYWD.V97` in the 1996 Edition

```
 2 James-Stein
 2 James-Stein estimation
58 James-Stein estimator
 1 James-Stein predictor
 1 Truncated James-Stein estimator
```

show that there are 64 database records with the expression "James-Stein" in Field 4. A search of the database shows that in addition to these records, an *additional* 41 records contain "James-Stein" in the title field (Field 2).

7.4 Index of Additional Searchable Terms

This short file, named `F6.V97`, is a tabulation of the semicolon-separated keywords/phrases in Field 6 of the database. The two fields, "count" and

"term", are separated by a blank; the count is right justified in a field of width 4.

In 1994 thousands of database records were "reTEXed" with regard to math mode. For example, "L2" (originally entered as `L2`) was changed to "L_2" (entered as `$L_2$`), Similarly, "t test" (`t test`) became "*t* test" (`$t$ test`). To simplify searches, common expressions such as "`$t$ test`" and "`$F$-statistic`" were placed in Field 6, stripped of `$`-signs and with blanks replaced by hyphens. Searching on the stripped version of one of these math fragments is guaranteed to retrieve all records containing the TEXed version(s).

For example, the following entries from `F6.V97` (1996 Edition)

```
239 t-test
  1 Two-sample t-test
```

shows that there are 240 database records with the term "t-test" in Field 6, and accordingly, either the term "*t* test" (`$t$ test`) or "*t*-test" (`$t$-test`) in Fields 2 or 4.

7.5 Inverted Indexes

The fastest search software makes use of "inverted indexes" to the database records. Most users will not need to know details of the structure or content of these indexes, but this section provides the details for those who wish to use the inverted indexes in their own programs to access the database.

Douglas Bates and Paul Tukey defined the format used for the Inverted Indexes, which were originally designed as part of their client-server search software for *CIS/ED* on Unix™-based computers. That search package (and its successors, provided on the CD-ROM and available from `Statlib` and the *CIS* web page (`http://www.stat.uchicago.edu/~cis`) includes a program to create an inverted index from the actual database files. The index for the entire *CIS/ED* is approximately 60% as large as the database itself. The inverted index makes it possible to "look up" words and to retrieve a list of all records in *CIS/ED* containing those words.

In *CIS/ED* there are two sets of inverted indexes, which reside in the directories `INDEX` and `INDEX2`, respectively. These index sets are identical in structure, and similar in content and size.

7.5.1 Inverted Indexes in INDEX

These indexes are used by several search programs, e.g., the Bates/Tukey software and CIScd. There is one *index file* for each of the data files in the database. Each such index file contains a list of the words contained in Fields 2 through 6 of the corresponding database file, together with a coded list of positions at which the word can be found. Here a "word" is defined to be a maximal length string of letters, digits, and underscore characters (_); upper-case letters are folded to lower-case prior to indexing. Words that don't begin with a letter (e.g., numbers and dates) are not indexed, and neither are one-character words and very common words (**and**, **the**, **of**, etc). Words exceeding six characters in length are treated as if they consisted of the *first six characters only.* Thus, **statistic**, **statistics**, **statistician**, and **statistical** would share the index entry **statis**.

Linking the separate index files together is a *master index*, which has an entry for each word appearing in at least one index file. These entries consist of the word, followed by a coded list. This list identifies each index file containing the word, the number of times the word appears in that index file, and the location of that word's entry in the index file. If a word appears only once in an index file, the master index records the position in the data file rather than the position of a list in the index file.

7.5.2 Inverted Indexes in INDEX2

The second set of inverted indexes is created in the same way as for **INDEX**, except that in the second set words exceeding six characters in length are *not* collapsed to their six-character root. Thus, in **INDEX2**, **statistic**, **statistics**, **statistician**, and **statistical** each have separate index entries. By inspecting the keys in the master index, you can easily tell which variants of a word you want to retrieve and which you do not. The indexes in **INDEX2** are used by the UCcis search software.

7.6 Thesaurus

Current Index to Statistics printed volumes contain lines suggesting items to "See..." or "See Also...." Some of these have to do with how the printed volumes are organized (e.g., time series is found under time, but not under series), and some offer suggestions as to alternative terminology (e.g., mark-recapture is related to capture-recapture; LISREL stands for

linear structural relations; clustering and classification are related ideas). The latter kind of information can also be helpful in searching *CIS/ED*.

This edition contains the file `SEEALSO.V97` which incorporates some cross-listings taken from printed CIS volumes with others that we thought would be particularly useful for computer searches. This file contains variable-length records, each containing a list of related words. The two or more items within a record are separated by various symbols: indicating transition from a general to a specific concept (`>`), full version of an acronym (`=`), various levels of similarity or parallelism (`, ; #`), or an antonym (`|`).

To improve the thesaurus, *CIS* solicits comments from users and contributed entries from experts in various fields.

8 Abstracts and Book Reviews

Over the long term, it is our hope to be able to include abstracts for all refereed journals affiliated with ASA and IMS, as well as abstracts for some other important publications.

Abstracts accompany entries for the following journals beginning with the publication years given in parentheses.

- *The American Statistician* (1996)
- *Journal of Agricultural, Biological, and Environmental Statistics* (1996)
- *Journal of the American Statistical Association* (1996)
- *Journal of Business and Economic Statistics* (1996)
- *Journal of Computational and Graphical Statistics* (1992)
- *Journal of Statistics Education* (1993)
- *Technometrics* (1996)

Book reviews from those journals in the list above that review books are also included with the corresponding entries in the database.

Entries in the data base for which abstracts or book reviews are available are denoted by a lower-case letter (`j` or `e`) in column 28 of Field 1. The detailed file structure for abstracts is described in the *CIS/ED Technical Reference Manual*.

Editors and publishers of journals who would like to have abstracts of their articles included in *CIS/ED* should contact Michael J. Wichura for instructions on preparing submissions.

Part IV
Coverage

9 The Gap Project

Bruce E. Trumbo, Associate Database Editor
California State University, Hayward
Brian M. Munneke, Assistant Database Editor
California State University, Hayward and Purdue University

9.1 Editorial policy

The Gap Project seeks to include in *CIS/ED* information on scientific papers that appeared in selected statistics and probability journals before 1975. The selection criteria for this project include: scholarly importance of the early papers appearing in the journal, availability of financial or editorial support from the publisher, the lack of systematic coverage in other bibliographic sources, and availability of the journal in libraries.

Abstracts, book reviews, news notices, and officer reports are generally not included in this project. Discussions of papers are sometimes indicated only by including the expression "(with discussion)" after the title of the main paper, rather than by making a separate record for the contribution of each discussant; in this case the pages given for the main paper include the discussion. As a general practice, records created under the Gap Project include the titles of non-English papers both in English translation and in the original language (inside parentheses).

Especially for papers published before 1965, key words may be present only for papers with particularly uninformative titles. Unless otherwise indicated below, key words were chosen by Bruce Trumbo. Since the publication of author's key words and abstracts with papers was not common before 1975, many of these key words are based on introductions, section headings, and summaries of papers. Some of the key words use terminology that came into use after the paper was published.

Some of the Gap Project records included in this edition of *CIS/ED* are still in draft form. These have been entered carefully from a variety of

sources, but have not yet been fully verified by comparison with title pages of the original articles. For many Gap Project journals, *CIS* records in the period 1975–79, which were previously reclaimed from deteriorated media, have been compared with original sources and brought up to current *CIS* standards. (Users interested in the details of the editorial status, editorial assistants, and copyright holders of records for articles published before 1980 may want to browse the administrative records that precede the bibliographic records for each volume of the journals involved.)

9.2 Contents

The GAP records comprise over 2Mb of material; it is estimated that the Gap Project has now captured about a quarter of the information that might have been included had the *Current Index to Statistics* been in existence in the period from 1965–74. In view of our guidelines for inclusion, we hope that this is, roughly speaking, the part of the bibliographic record from that period that will be the most important for current users.

Previous editions of *CIS/ED* contained information on pre-1975 publications in the journals listed in order of inclusion in items 1–6 below. Dates in *italics* correspond to the initial volume of a journal.

1. As part of the preparation of Trumbo and Burdick [1], IMS provided financial support for data entry and editorial assistance in covering:

   Annals of Mathematical Statistics (1960–72)
   Annals of Probability (*1973*–74)
   Annals of Statistics (*1973*–74).

2. As part of the preparation of Trumbo, Gbur, and Suess [2], ASA provided financial support for data entry and editorial assistance in covering:

   Journal of the American Statistical Association (1955–92).

3. The Technometrics Management Committee provided financial support for data entry and full editorial cooperation in preparing records for:

   Technometrics (*1959*–74).

William Q. Meeker organized a team of 27 associate editors and frequent contributors to *Technometrics* to assist with the selection of key words for these volumes.

4. The International Biometric Society provided financial support and Klaus Hinkelmann and Alexandra Kapatou provided editorial support (including assistance in selecting key words) in preparing records for:

   Biometrics (1965–74).

5. The Biometrika Trust supplied back issues to assist in the preparation of records for:

   Biometrika (1965–74).

6. Using CIS funds, draft coverage (in 1993) and full coverage (1994) was included in *CIS/ED* for:

   Applied Statistics (1965–74).

   For the entire period 1965–92, the numbers associated with computational algorithms and remarks on algorithms have been included, and appropriate cross referencing of algorithms and associated remarks has been completed.

7. Gap coverage for the journals listed below was completed in 1994:

   Advances in Applied Probability (*1969*–74)
   The American Statistician (*1947*–74)
   Australian Journal of Statistics (*1959*–74)
   Canadian Journal of Statistics (*1973*–74)
   Journal of Applied Probability (*1964*–74)
   Journal of the Royal Statistical Society, Series A (1965–74)
   Journal of the Royal Statistical Society, Series B (1965–74)
   Sankhyā, Series A (1965–74)
   Sankhyā, Series B (1965–74)
   Zeitschrift für Wahrscheinlichkeitstheorie (*1962*–75).

   Data entry for the *The American Statistician* was financially supported by ASA in preparation for a forthcoming index on diskette of that journal [3]; all other data entry was done using CIS funds.

An excellent index prepared by the Applied Probability Trust covering the years to be added to *CIS/ED*, greatly assisted in preparing records for *Journal of Applied Probability* and *Advances in Applied Probability*. We thank Imke Janssen for her assistance with the translations of the many German titles in *ZfW*, and G. P. H. Styan for supplying back issues of the *Canadian Journal of Statistics* and for other assistance.

8. Coverage for the journals listed below was completed in 1995:

   Econometrika (1965–74)
   Theory of Probability and its Applications (*1956*–79)

   For the latter journal, *CIS* has indexed the English translation. However the Russian original was used to correct the large number of faulty transliterations of Russian names that appeared in the English edition.

9.3 General Acknowledgments

Almost all of the data entry for the Gap Project has been done by Bill Brent, whose accuracy, efficiency, and patience have made a major contribution to the project. Thanks are also due to former students at California State University, Hayward, who have assisted in a variety of ways including proofreading and the attachment of correction notes to referent papers: Rima Bannerjee, Jeffrey Hill, Huu-Manh Hoang, Pamela D. Ochoa, Jason H. Stover, Eric A. Suess, Douglas Whitehouse.

References

[1] Trumbo, Bruce E, and Burdick, Richard. (1991). *Index of IMS Scientific Journals (1960–1989)*, Institute of Mathematical Statistics.

[2] Trumbo, Bruce E, Gbur, Edward E, Jr, and Suess, Eric A. (1993). *Journal of the American Statistical Association, Subject and Author Index (1955–1991)*, American Statistical Association.

[3] Trumbo, Bruce E, Munneke, Brian M, and Wichura, Michael J. (1996). *Index of The American Statistician (1947–1994)*, Computer searchable database with search software. American Statistical Association. [In preparation, expected early 1996].

10 License Information and Legal Matters

CIS/ED is distributed for the *Current Index to Statistics* by the IMS Business Office, 3401 Investment Blvd, Suite 7, Hayward, CA 94545 USA and the American Statistical Association, 1429 Duke Street, Alexandria, VA 22314 USA. Information about ordering subsequent editions of *CIS/ED* will be announced in ASA and IMS journals and may also be obtained from the offices above.

Search software supplied on the CD-ROM version is copyright by the authors of the software, who make it available to users of the database. The Current Index itself does not provide support for search software distributed with the database.

CIS/ED and this *User Guide* are protected under U.S. Copyright Law, Copyright © 1997, all rights reserved. Use of the database and *User Guide* is subject to a License Agreement with each site or individual member of ASA or IMS authorized to use them. In 1997, there is a single License Agreement with several licensing levels, one of which applies to your usage of *CIS/ED*. Below we give brief descriptions of the different levels, but you should carefully read the version of the License Agreement (supplied in the `LICENSES` directories on the CD-ROM) that applies to you because the informal language used here does not substitute for or modify the legal language of the License Agreement.

Fees to current licensees for subsequent versions are set considerably below the license fee charged to new users. Users are encouraged to license each consecutive version. Licenses of non-consecutive versions are treated as new licenses. Each annual version contains a complete, updated, and corrected copy of the latest edition of the database, which must be substituted for the earlier edition.

General Site License

We require that a site consist of a single contiguous geographic location, unless otherwise agreed in writing. Examples of a site might be a single office of an industrial division or government agency, or a university campus. *CIS/ED* may be installed on individual computers and/or on a network for on-site use by employees, students, and visiting scientists. There is no limit on the number of stations at the site or the number of simultaneous users at the site. However, off-site access via modems or networks is not permitted. Administrative, hardware, and/or software controls should be in place to prevent unauthorized transfer of the *CIS/ED* (or parts of it) off-site.

University Campus License

CIS/ED is licensed for use by faculty, staff, students, and authorized visiting scientists holding currently valid passwords for access to the computer system or network of a university campus. *CIS/ED* may be accessed from on or off campus. Administrative controls on passwords, their use, and their renewal should be in place to prevent *CIS/ED* (or parts of it) from being transferred to or used by those who are not currently part of the local campus community. Available only to educational institutions.

Commercial License

CIS/ED is licensed for use on an unlimited number of computers at a single site, as well as by staff holding currently valid passwords for access to a single computer system or network of a corporate entity. *CIS/ED* may be accessed either locally or remotely. Administrative controls on passwords, their use, and their renewal must be instituted to prevent *CIS/ED* (or parts of it) from being transferred to or used by those who are not currently employed by the licensee.

Four-station License

Access to the *CIS/ED* must be limited to four designated stations (either stand-alone or in a local area network). You may NOT make the *CIS/ED* available over a network that serves more than four stations, even if network software limits the number of simultaneous users. Remote access is not permitted. Not available to national government agencies.

Personal License

An individual is licensed as the sole user of *CIS/ED*, which may be installed on up to three computers under the licensee's control.

Users of the *CIS/ED* should have easy access to this *User Guide* and should be encouraged to read it. Except for Personal Licenses, a copy of the applicable License Agreement must be available to any station that can access *CIS/ED*.

In particular, all users must be informed that indexes, bibliographies, and other works based on more than 250 records from *CIS/ED* may not be published or distributed without written permission of the CIS Management Committee, and that published works prepared with the substantial assistance of *CIS/ED* should acknowledge its use and cite the 1997 edition as a reference. The following is a suggested citation:

> "Michael J. Wichura, Editor (1997). Current Index to Statistics Extended Database. American Statistical Association (Alexandria, VA) and Institute of Mathematical Statistics (Hayward, CA)."

Administrative records appear in many of the database files. These have various uses, which may differ from one edition to another: for example, they may carry editorial records and copyright notices. It is unlikely that ordinary searches of the database will retrieve any of them; if they are retrieved, their nature should be obvious and they are not likely to be confused with bibliographic records.

11 History of the *CIS/ED*

11.1 Information from printed volumes of CIS

The usefulness of this edition of *CIS/ED* depends in part on the efforts of past editors of the *Current Index to Statistics* printed volumes from whose files major portions of it were developed: Brian Joiner, Volumes 1–5 (covering years 1975–1979); James Gentle, Volumes 6–10 (1980–1984); Richard Burdick, Volumes 11–15 (1985–1989); and Edward Gbur, Volumes 16–20 (1990–1994). The work of their contributing editors, editorial collaborators, and editorial assistants, acknowledged in each printed volume of the *Current Index to Statistics*, was essential to preparing the original *CIS/ED* databases.

Beginning with the 1996 edition of *CIS/ED* and Volume 21 of the print version, data collection and editorial efforts for the electronic and print editions have been integrated. The new records on the CD-ROM and in the print volume are now produced from a single set of edited records.

Items considered for indexing in *CIS* include journal articles, books and edited books, articles from edited books and conference proceedings (provided that the proceedings are available to non-conference participants), book and software reviews, prepared discussions of articles, and correction notes. Items such as notices and advertisements, editorials, letters to the editor, and meeting announcements are not knowingly indexed.

Subject to the above proviso, *CIS* includes all articles from approximately 100 core journals, and selected articles from approximately 400 non-core journals and other sources. The list of core journals and the extent of their coverage in *CIS/ED* and Volume 22 for 1997 is in the file `CORE.V97` in the `ABBREV` directory on the CD-ROM.

The "Tukey indexes" to the statistical literature (available from the American Mathematical Society both in printed form and through some of the MathSci on-line services offered by AMS) cover various important

statistical journals for differing periods of time, mainly ending in the mid-1960s. Because the *Current Index to Statistics* was not established until 1975, there was a gap of about ten years in the systematic bibliographic coverage of the statistical literature. Since its inception, an important goal for the CIS computer-readable database has been to provide useful coverage for the period of this gap. This effort has had the official backing of CIS, ASA, and IMS. Chapter 9 details the current status of the Gap Project, under the editorship of Bruce Trumbo, and acknowledges the efforts of those who have made particular contributions to it.

11.2 Acknowledgements

In addition to everyone involved with *CIS* since its inception, several more people have made contributions of time, energy, and expertise to improve various aspects of the *CIS/ED* and deserve acknowledgement and thanks. Bruce Trumbo has been indefatigable in expanding coverage, and he has also made major contributions to the planning and implementation of the many changes and improvements from the inception of the database to the present. Michael Wichura has devoted many late nights and early mornings to the task of making the *CIS/ED* a cohesive work of the highest quality and consistency. Mike is responsible for literally thousands of changes, corrections, improvements and innovations. Ed Gbur has made major improvements to the *CIS* abbreviation files, and he has generously provided valuable advice to the current editors. Bill Meeker wins first prize for bringing to our attention the largest number of typos, inaccuracies, and other infelicities in our early editions. Correcting these has often led to identifying and systematically correcting large numbers of similar errors. Finally, Paul Shaman and the other members of the Management Committee have done an excellent job of making the publication responsive to the evolving needs of the statistics community.

12 Errors, Omissions, and Corrections

Thousands of errors have been found and corrected since the first edition of this database. Please report any errors you find. No error is too trivial to report.

To enable us to correct any errors you discover, please provide us with a printout of the defective record with the error clearly marked. We will try to compare suggested corrections against the original source. Where possible, please send a photocopy of the title page or first and last pages of the item in question (marked with publisher, journal name, date, volume, inclusive pages, etc., as appropriate) to show how the *CIS/ED* record should read. Corrections can be faxed to the *CIS/ED* Editor at (773) 702-9810, by e-mail to `cised@galton.uchicago.edu`, or by postal mail to *Current Index to Statistics,* Department of Statistics, The University of Chicago, 5734 University Avenue, Chicago, IL 60637 USA.

Photocopies of source material are *required* for suggested additions to *CIS/ED*. The *Current Index to Statistics* does not index all articles with statistical content. We welcome lists of possible omissions. However, they will be included in future editions only if they meet the current criteria for indexing *and* the necessary documentation is included. These materials can be sent to Klaus Hinkelmann, *Current Index to Statistics,* Department of Statistics, Virginia Tech, Blacksburg, VA 24061-0439 USA.